REMARKS

Claims 1-17, all the claims pending in the application, stand rejected on prior art grounds. Applicants respectfully traverse these rejections based on the following discussion.

I. The Prior Art Rejections

Claims 1, 6, and 11 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Kostoff et al., hereinafter "Kostoff" (U.S. Patent No. 5,440,481). Claims 2-5, 7-10, and 12-17 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Kostoff and in further view of Kirsch et al., hereinafter "Kirsch"(U.S. Patent No. 6,070,158), Kobayashi (U.S. Patent No. 5,742,834) and Turney (U.S. Patent No. 6,470,307). Applicants respectfully traverse these rejections based on the following discussion.

A. The Rejection Based on Kostoff

In response to previous arguments, the Office Action succinctly states (on page 14, section 6) that the primary difference between the claimed invention and Kostoff is that Kostoff does not discard the unused portion of the dictionary; and, the Office Action argues that it would be obvious to reduce the size of the dictionary in order to fit the dictionary within a specific memory constraint. However, a more important difference between the claimed invention and Kostoff is that the size of the dictionary is limited before the frequency of phrases in the document that contain words in the dictionary is determined. This is important because the number of phrases grows exponentially with the size of the corpus. Therefore, by reducing the size of the dictionary before determining the frequency of phrases containing words in the dictionary, the claimed invention produces exponential gains in processing speed and memory usage. In other words, the claimed invention involves more than just reducing the dictionary to meet a memory constraint. In the claimed invention, the dictionary is reduced in order to

substantially simplify the subsequent process of determining the frequency of phrases in the document containing words in the dictionary.

The claimed invention first limits the dictionary to only the top number of most frequently occurring words and then only considers phrases that contain these words. The invention avoids maintaining a list of all potential phrases in the text corpus. The problem with maintaining all potential phrases is that the number of phrases grows exponentially with the size of the corpus. The invention avoids this problem by fixing the size of the dictionary up front (user specified maximum dictionary size, M), then finding the M most frequent words and then only creating phrases using these M most frequent words. To the contrary, the Kostoff patent creates a list of all words and N-word phrases sorted by frequency. This is not practical for a large text corpus since such a list would be too large for most computer memory to hold.

The Office Action admits that Kostoff does not explicitly teach the claimed process of limiting the number of words that are used to establish the most frequently occurring phrases by limiting the dictionary size, but the Office Action argues that such a feature would have been obvious. More specifically, the Office Action notes that Kostoff describes that the size of the list of trivial phrases is limited by memory constraints (col. 4, lines 42-45) and that the number of phrases output to the user can be limited to those having high user interest, such as the top 60 most frequent phrases (col. 5, line 59-col. 6, line 64). Then, the Office Action argues that this would motivate one to limit the dictionary size to accommodate for hardware memory constraints.

Applicants respectfully disagree with this logical argument of obviousness for a number of reasons, including the fact that Kostoff requires that the dictionary must include all words in the documents (except for the trivial phrases mentioned above). More specifically, Figure 2 and col. 4, lines 52-55 states that the system and methodology in Kostoff "is required to use the entire full-text database to create lists of phrases." Therefore, Applicants submit that Kostoff directly teaches away from the claimed limitation that explicitly does not use all the words from the documents, and instead limits the dictionary to only the number of most frequently occurring words that will fit into the limited size dictionary. When a reference teaches away from the

claimed invention it actually demonstrates that the claimed invention is not obvious.

Thus, in a first respect, since Kostoff "is required to use the entire full-text database to create lists of phrases" it cannot teach or suggest "creating a dictionary of most frequently occurring words in said documents as limited by said maximum dictionary size, such that said dictionary contains less than all words in said documents" as defined by independent claims 1, 6, and 11. This requirement in Kostoff teaches away from the claimed invention and, therefore, Kostoff cannot teach or suggest this feature.

Further, the manner in which Kostoff would deal with memory and other limitations is conceptually different than the claimed invention. For example, in order to deal with memory constraints, Kostoff creates a list of trivial phrases that can be excluded from analysis (col 4, lines 39-49). This is essentially a fixed list in Kostoff that may or may not be effective in limiting the memory usage. To the contrary, the claimed invention limits the size of the dictionary, thereby providing for a more consistent and precise control of memory usage. In addition, the processing in Kostoff always uses all words in the database (except trivial words) and merely limits the number of phrases that are output (col. 5, line 59-col. 6, line 64). Thus, since all words are used in the most frequent phrase processing of Kostoff, no memory is conserved. To the contrary, the claimed invention first limits the dictionary to only the top number of most frequently occurring words and then only considers phrases that contain these words.

As explained on page 4, lines 4-9 of the application, the invention allows the user to specify the size of the vector space model to be used in text clustering of a document corpus, as well as the maximum number of words that can occur in a phrase. The invention will find all of the phrases, up to the user specified length, that occur with the greatest frequency. The total number of phrases returned will depend upon the user specified maximum dictionary size.

One distinction of the invention when compared to Kostoff is that the invention avoids maintaining a list of all potential phrases in the text corpus. The problem with maintaining all potential phrases is that the number of phrases grows exponentially with the size of the corpus. The invention avoids this problem by fixing the size of the dictionary up front (user specified

maximum dictionary size, M), then finding the M most frequent words and then only creating phrases using these M most frequent words. To the contrary, the Kostoff patent creates a list of all words and N-word phrases sorted by frequency. This is not practical for a large text corpus since such a list would be too large for most computer memory to hold.

Therefore, it is Applicants' position that Kostoff does not teach or suggest "creating a dictionary of most frequently occurring words in said documents as limited by said maximum dictionary size, such that said dictionary contains less than all words in said documents . . . wherein said dictionary size limits the number of words and phrases maintained in said dictionary" as defined by independent claims 1 and 11 and similarly defined by independent claim 6. Previous methodologies that have suggested a lexical phrase generation technique have not described the space and time efficient implementation for discovering such phrases that the invention utilizes. The invention's implementation is designed to quickly find a maximal frequency term dictionary of a given size using the smallest possible amount of memory.

Therefore, because the prior art of record does not teach or suggest the claimed invention, Applicants respectfully submit that independent claims 1, 6, and 11 are patentable over the prior art of record. In view the foregoing, the Examiner is respectfully requested to reconsider and withdraw this rejection.

B. The Rejection Based on Kostoff in view of Kirsch and further in view of Kobayashi and Turney

With respect to dependent claims 2-5, 7-10, and 12-17, the Office Action makes reference to the prior art Kirsch, Kobayashi, and Turney as teaching concepts such as removing punctuation, replacing words with synonyms, removing stop words, removing duplicates words, clustering, etc. Therefore, the additional prior art references are not utilized to teach or suggest (and do not teach or suggest) the claimed features defined by independent claims 1, 6, and 11. Therefore, it is Applicant' position that the proposed combination of all references still does not teach or suggest "creating a dictionary of most frequently occurring words in said documents as

limited by said maximum dictionary size, such that said dictionary contains less than all words in said documents . . . wherein said dictionary size limits the number of words and phrases maintained in said dictionary" as defined by independent claims 1 and 11 and similarly defined by independent claim 6. Therefore, it is Applicants position that none of the prior art of record teach or suggest the invention defined by independent claims 1, 6, and 11 and that such independent claims are patentable over the prior art record.

Further, dependent claims 2-5, 7-10, and 12-17 are similarly patentable, not only by virtue of their dependency from a patentable independent claim, but also by virtue of the additional features of the invention they define. Therefore, Applicants submit that dependent claims 2-5, 7-10, and 12-17 are patentable over the prior art of record and respectfully request that the Examiner reconsider and withdraw this rejection.

II. Formal Matters and Conclusion

In view of the foregoing, Applicants submit that claims 1-17, all the claims presently pending in the application, are patentably distinct from the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.

Should the Examiner find the application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary.

09/629,831

Please charge any deficiencies and credit any overpayments to Attorney's Deposit Account Number 09-0441.

Respectfully submitted,

Dated: 4-4-05

Frederick W. Gibb, III
Registration No. 37,629

McGinn & Gibb, PLLC
2568-A Riva Road, Suite 304
Annapolis, MD 21401
301-261-8071
Customer Number: 29154